

COVID-19 Case Surveillance Public Use Data Utility Summary

Users should consider the level of completeness, including suppression levels when planning their analyses and use of public datasets. Privacy protections will suppress field values to reduce reidentification risks. Completeness varies by jurisdiction (i.e., state, local, and territorial) and time period. Variables are consistently coded to the value “Unknown” when jurisdictions specify in the case data submitted to CDC that the value is unknown, the value “Missing” when jurisdictions do not provide a value, and the value “NA” when the value is suppressed as part of privacy protections.

Dataset version: 2021-06-21

Total records in dataset: 26,887,803

Quick Summary

	all_fields	quasi_fields
total_fields	12	3
total_records	26,887,803	26,887,803
total_cells	322,653,636	80,663,409
missing_fields	66,043,098	890,899
missing_pct	20%	1%
complete_fields	256,610,538	79,772,510
complete_pct	80%	99%
unknown_fields	22,471,857	9,526,470
unknown_pct	7%	12%
suppressed_fields	106	106
suppressed_pct	0%	0%
available_fields	234,138,575	70,245,934
available_pct	73%	87%

Field-level Utility Summary

	suppressed	suppressed_percent	missing	missing_percent
race_ethnicity_combined	16	0.0%	652,501	2.4%
sex	22	0.0%	54,652	0.2%
age_group	68	0.0%	183,746	0.7%
records_with_any_field	68	0.0%	830,197	3.1%